The Mechanics and Physics of Defect Nucleation

Ju Li

Abstract

The following article is based on the Outstanding Young Investigator Award presentation given by Ju Li on April 19, 2006, at the Materials Research Society Spring Meeting in San Francisco. Li received the award "for his innovative work on the atomistic and first-principles modeling of nanoindentation and ideal strength in revealing the genesis of materials deformation and fracture."

Defect nucleation plays a critical role in the mechanical behavior of materials, especially if the system size is reduced to the submicron scale. At the most fundamental level, defect nucleation is controlled by bond breaking and reformation events, driven typically by mechanical strain and electronegativity differences. For these processes, atomistic and first-principles calculations are uniquely suited to provide an unprecedented level of mechanistic detail. Several connecting threads incorporating notions in continuum mechanics and explicit knowledge of the interatomic energy landscape can be identified, such as homogeneous versus heterogeneous nucleation, cleavage versus shear-faulting tendencies, chemomechanical coupling, and the fact that defects are singularities at the continuum level but regularized at the atomic scale. Examples are chosen from nanoindentation, crack-tip processes, and grain-boundary processes. In addition to the capacity of simulations to identify candidate mechanisms, the computed athermal strength, activation energy, and activation volume can be compared quantitatively with experiments to define the fundamental properties of defects in solids.

Small Length Scale and High Strength

It is well known to materials scientists that material behaviors, especially mechanical properties, are controlled by defects. Frenkel estimated the ideal strength of a crystal to be around one-tenth of its modulus. But most metallic materials we exploit today deform at around one-thousandth of their moduli, when dislocations start to move. For brittle materials, Griffith and Weibull delineated the sabotaging effect of cleavage defects in relation to their size and population. It turns out that conventional monolithic ceramics perform even much worse than metals in tension.

The gap of 10^{-3} to 10^{-1} just described—a factor of 100 below the ideal strength—is actually good news: it means we still have plenty of room for improvement. The most effective way to potentially close the gap significantly is to change the length scale of the material. This is demonstrated by recent experiments[1,2] (Figure 1a) on focused-ion-beam-carved Au nanopillars. As the pillar diameter approached hundreds of nanometers, one measured a uniaxial compressive strength of 800 MPa (Figure 1b).[2,3] This is extremely high for Au. Density functional theory[4] (DFT) calculation predicts an ideal shear strength of 850 MPa to 1.4 GPa for Au, depending on the loading constraints.[5] One needs to further consider the weakening effect of free surfaces. In this case, it appears the ideal *surface* strength,[6,7] rather than the ideal *bulk* strength,[5,8] should be used as the benchmark. Preliminary molecular dynamics (MD) calculations using embedded-atom potentials[9] suggest the ideal surface strength of metals to be around half of its ideal bulk strength, as atoms on the surface are weakened due to the missing neighbors. The experimental value is thus near the ceiling of what theory expects it can be. The achievability of such high strength experimentally is confirmed by independent measurements on nanoporous Au,[10,11] where individual metal ligaments are tens to hundreds of nanometers in thickness.

Experimental studies on high-strength materials systems and phenomena (defined as sustaining stress broadly and persistently at a significant fraction of the ideal strength) have blossomed in recent years. One reason is the refinement of nanoindentation[12–16] and other nanoscale mechanical experiments, which allows one to study near-ideal strength behavior quantitatively in a controlled fashion. In the case of nanoindentation, the small length scale is not that of the tested material, but the extent of high stress under the indenter, characterized by the size of the contact zone. Since elasticity is governed by the same family of equations as electromagnetics, a spherical indenter tip works somewhat like a lens, "projecting" the applied force to "focus" the maximum shear stress to an internal point inside the sample, away from the surface. This gives one a chance to probe bulk properties[17] near the high-stress spot, which otherwise could be dominated entirely by the surface. Table I shows some recent experimental results. In the second column, we have the maximum shear stress inferred from linear elasticity at major displacement bursts[13,14,18] (serrated plastic flow) during nanoindentation. These should be compared with the rightmost column, which is ideal shear strength at zero confining pressure calculated from density functional theory.[5] It is somewhat intriguing to see that the experimentally inferred shear strength from nanoindentation correlates with the ideal bulk strength obtained from quantum mechanical theory, having at least similar *magnitudes*.

What do these new experimental results say about defects? I would suggest that as materials and devices shrink in size, defect nucleation plays a more important role in defining their behavior. So in these high-strength systems and phenomena, we need to pay extra attention to defect nucleation.[19–21]

Defect behavior can be roughly categorized into nucleation and mobility. The nucleation stage is when a defect forms its individual, independent identity. Mobility is about how the defect translates without changing its identity. The Frank–Read dislocation source is a classic example of nucleation, because a free, separate dislocation loop forms each time the source operates. The athermal stress (defined later) required to give birth to a new loop is $\tau_{ath} = G\mathbf{b}/d$, where G is the shear modulus, $\mathbf{b}$ is the Burgers vector, and d is the distance between two pinning points.

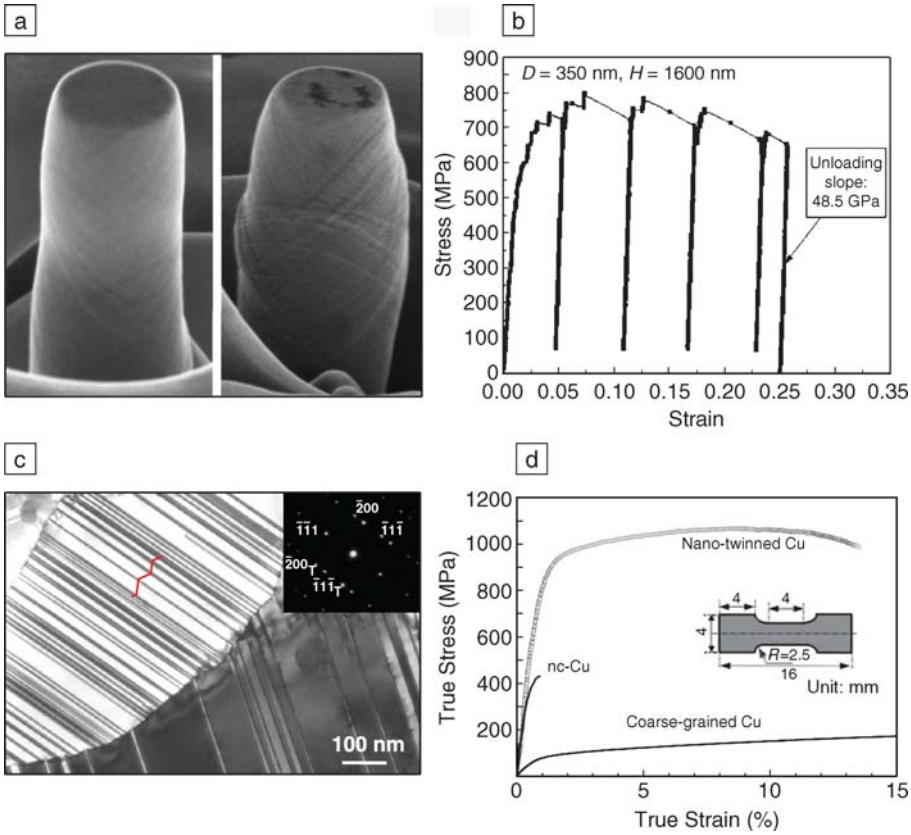

Figure 1. (a) Focused-ion-beam-carved single-crystal Au nanopillar (diameter, 660 nm), after compression by a flat punch (left, ~15% strain; right, ~30% strain). (b) Uniaxial compression stress–strain curve for a 350-nm-diameter pillar. (Adapted from Greer et al.[2]) (c) Nano-twinned Cu synthesized by pulsed electrodeposition. The red zigzag line (added by the author) indicates the necessary path for dislocation motion. (d) Tensile stress–strain curve of nano-twinned Cu compared with nanocrystalline (nc) Cu with only general grain boundaries and coarse-grained Cu, showing high strength and good ductility. Inset shows shape and size of sample used for tensile testing. (Adapted from Lu et al.[34])

Table I: Recent Experimental Results of Maximum Shear Stress τ_{max} Sustained at a High-Stress Spot inside a Tested Sample during Nanoindentation.

Tested Sample	Maximum Shear Stress[a]	Experimental Reference	Ideal Shear Strength at Zero Pressure
Pt	4.4 GPa	Mason et al.[23]	5.3 GPa[24]
Al	~2 GPa	Minor et al.[15]	2.8 GPa[5,8]
Cu	~6 GPa	Wang and Lu[25]	2.2 GPa[5,8]
Ni_3Al	~8 GPa	Wo et al.[26]	11 GPa ($G/2\pi$)
Au	2–6.5 GPa	Asenjo et al.[27]	850 MPa[5]
Ni	~8 GPa	Lorenz et al.[28]	5 GPa[5]
W	~15 GPa	Lorenz et al.[28]	17 GPa[5]
SiC	~25 GPa	Schuh and Lund[29]	31 GPa[5]
Cr_3Si	18.1–21.7 GPa	Bei et al.[30]	20.5 GPa ($G/2\pi$)
GaAs	5–10 GPa	Leipner et al.[28,31]	8.6 GPa ($G/2\pi$)

[a] Maximum shear stress inferred from linear elasticity $\tau_{max} \approx 0.31(6PE^{*2}/\pi^3R^2)^{1/3}$ during nanoindentation. P is indentation load, R is radius of curvature of the indenter tip, E^* is the effective elastic modulus, and G is the resolved shear modulus.

Now in general, d is some kind of length scale—for example, the average grain size—and this inverse correlation between strength and length scale happens frequently in materials (the exponent[22] may not be -1; it may be $-1/2$ or some other value). The Frank–Read source thus has a sensitive length-scale dependence. Subsequent expansion of this loop against, for instance, solute drag, is clearly an issue of mobility.

The distinction between nucleation and mobility is not always clear-cut. For example, forest dislocation hardening,[32] which has similar length scaling as operating the Frank–Read source, is probably better considered a problem of mobility only because we envision the advancing mobile dislocation maintaining its identity as it overcomes obstacles (i.e., the "forest" of dislocations it encounters in other slip systems). It is perhaps useful then to acknowledge there is a continuum between nucleation and mobility, depending on the degree of identity change as mobile defects carry out free energy relaxation.

With that in mind, let us examine another high-strength system, nanocrystals.[33] Figure 1c shows high-purity Cu with nanoscale growth twins synthesized by pulsed electrodeposition, with twin lamella thickness on the order of 10 nm.[34] This material has a tensile strength approaching 1 GPa, again a respectable fraction of copper's ideal strength,[5,8] although a lesser fraction compared with previous examples. From strain-rate sensitivity measurements,[35] the activation volume[32] was determined to be $12\mathbf{b}^3$ to $22\mathbf{b}^3$, which suggests interface-mediated slip transfer reactions to be the rate-limiting process.[36,37] To carry out plastic deformation, dislocations are forced to execute the kind of zigzag motion illustrated in Figure 1c, because the active slip system has to change from one lamella to the next. One may alternatively say new dislocations (either interfacial dislocations or bulk dislocations[37]) must be nucleated at each twin interface in order to satisfy Burgers vector conservation, and these nucleation barriers are the reason for the material's extraordinary strength. Incidentally, this and some other nanocrystalline systems also manifest reasonably good tensile ductility (Figure 1d),[38,39] bolstering the hope that high-strength nanocrystalline systems may be developed into structural materials that are truly superior to conventional coarse-grained materials. Other high-strength systems include metallic glasses,[40,41] materials under shock loading,[42] and carbon nanotubes.

Understanding defect nucleation in stressed materials is not just important for traditional structural applications. Strained

silicon is a new technology[43,44] where a rather significant stress (for brittle materials, if they are macroscopic) is applied intentionally to nanoscale regions of the integrated circuit to improve carrier mobilities. This technology is now functioning in many of our personal computers whose chip feature size has already reached 65 nm, and is regarded as one of the major means to delay the breakdown of Moore's law. Eventually, this technology may be limited by defect generation,[44,45] which relaxes the strain. Similarly, how defects form in carbon nanotubes[46–49] may be important for the assembly and functioning of nanotube-based devices. As small-scale materials have increasingly become the basic building blocks of our technology, and as there is clear experimental evidence that small-scale materials possess fundamentally different mechanical properties from their macroscale counterparts, it has become imperative to understand defect nucleation at a more fundamental level.

Atomistic Modeling and Mechanics Framework

While defect nucleation is of most importance in high-strength systems, it often cannot be ignored in low-strength systems. Dislocation nucleation near crack tips has long been suspected to be a controlling factor in the brittle-to-ductile transition.[50–54] Dislocation cross-slip,[55] a critical process in the thermal recovery of coarse-grained materials, may be regarded as a nucleation event because of the change in dislocation character (slip plane). Finally, the mobility of a larger defect is often controlled by the nucleation of smaller defects on itself, for instance, double kinks in the thermally activated motion of screw dislocations in semiconductors,[56,57] bcc metals,[58] and crack kinks in the lattice trapping of cleavage cracks.[59,60] To rationalize and sometimes fit experimental results, a semi-analytical framework has been developed historically[32,61,62] on the basis of transition-state theory and the Peierls concept.[50,51,63] It is only in recent years, however, that this analytical framework has been confronted with data from direct atomistic calculations.[64] Such computer models typically involve 10^4–10^6 atoms, as the 3D geometry tends to be complex and there is often a significant elastic interaction component. For this number of atoms, high-accuracy first-principles calculations cannot be applied to attack the problems directly, and employment of empirical interatomic potentials[9] is necessary. Novel transition-pathway search algorithms such as the nudged elastic band (NEB) method[65] are used to explore the potential energy landscape and find the saddle point of such nanomechanical thermally activated processes.[54–56,58,60,66]

Atomistic calculations are uniquely suited for studying unit processes, because one has perfect control of the initial and boundary conditions in the calculation and access to all energetic and geometric information. Calculations employing empirical interatomic potentials are limited by the accuracy of the potential. But this should not be its Achilles' heel in this context, because the basic mechanics of the framework should apply to the real systems as well as to "model" systems in the computer. It is therefore critical to phrase the atomistic calculation results in the language of the mechanics framework, to identify the important material parameters (surface energy, unstable stacking energy,[63] unstable twinning energy,[67] multiplane generalized stacking fault energy,[8,68] shearability,[5] etc.) for the particular atomistic process through, for instance, parametric studies. The best outcome of an atomistic calculation is to improve a mechanics framework by introducing new materials concepts and/or improving old connections. This forces one to throw out less relevant details (model reduction). Currently, real materials are often still too complex chemically for direct atomistic calculations with empirical potentials. But, if we have established a mechanics framework that is well calibrated against atomistic calculations for simpler systems, most of the critical material parameters for these chemically complex systems may be obtainable from *ab initio* calculations,[69] circumventing the need for direct atomistic calculations case by case.

The usual starting point for discussing defect nucleation is the harmonic transition-state theory (TST) expression $R = \nu\exp(-Q/k_BT)$, where R is the nucleation rate, Q the activation free energy, k_BT the amount of thermal fluctuation, and ν the attempt frequency. While harmonic TST can fail, for many problems it is a sufficiently reasonable approximation. At low T, Q is dominated by the potential energy difference $U^{\text{saddle}} - U^{\text{initial}}$ between the rate-limiting saddle point and the initial local minimum in the $3N$-dimensional configurational space $\mathbf{x}^{3N}$, where N is the total number of atoms. The minimum energy path (MEP), also called the reaction or transition path, is a parameterized curve $\mathbf{x}^{3N}(s)$ that connects two local minima on the potential energy landscape $U(\mathbf{x}^{3N})$ via one or more saddle points; s is called the reaction coordinate, often taken to be the hyperspace arc length $\int(d\mathbf{x}^{3N}\cdot d\mathbf{x}^{3N})^{1/2}$. When T is large and the entropy effect becomes significant, Q is the free-energy difference between the $3N-1$ dimensional dividing surface, leading to the final state and the initial potential energy basin,

$$Q = U^{\text{saddle}} - U^{\text{initial}} - k_BT\ln\left(\prod_{i=1}^{3N-1}\nu_i^{\text{initial}}\Big/\prod_{i=1}^{3N-1}\nu_i^{\text{saddle}}\right), \quad (1)$$

where ν_i is the frequency of a vibrational normal mode, with the normal mode along the MEP excluded from the quotient. Both the MEP and Q depend on the local stress σ: $Q = Q(\sigma)$; or in the case of crack-tip activation, the stress intensity factor K: $Q = Q(K)$.

Heterogeneous defect nucleation means the birth of a new defect, which at one point during its nucleation must be topologically connected to a preexisting defect. The new defect should have a distinct character from the preexisting defect, for instance, an embryonic dislocation emitting from a crack, or a daughter crack emitting from a mother crack on a different plane; otherwise this would simply become the propagation of the old defect. Homogeneous defect nucleation means birth of a new defect without the "direct aid" (topological contact) of preexisting defects; "indirect aid" via long-ranged elastic stress is allowed. However, its effect can then be absorbed into the local stress σ.

One may also find in some literature that the term "homogeneous nucleation" means nucleation of a symmetry-breaking defect from an old defect that had higher symmetry, for instance, nucleation of double kinks on a perfectly straight dislocation or nucleation of an embryonic dislocation loop out of an atomically flat surface. There are some merits to this definition. As long as the context is clear, there is no confusion. Here, I take homogeneous nucleation to mean defect nucleation within a perfect crystal, under high stress.

Athermal Threshold

For both homogeneous and heterogeneous nucleation, we define the athermal threshold stress σ_{ath} as when $Q(\sigma_{ath}) \equiv 0$ (see Figure 2). Since stress is a tensor, σ_{ath} is generally a five-dimensional "yield surface" in stress space. The athermal threshold condition corresponds to the limit where the driving force is so large that the activation barrier vanishes, so nucleation can proceed even at 0 K. For crack-tip nucleation, we can define the athermal threshold stress intensity factor[63] as when $Q(K_{ath}) \equiv 0$. Since there are three crack modes, K_{ath} is generally a two-dimensional surface. Note that K_{ath} is a different physical entity from the critical stress intensity factor K_c for brittle fracture in the well-known Griffith criterion, which is a thermodynamic condition for crack propagation. The Griffith

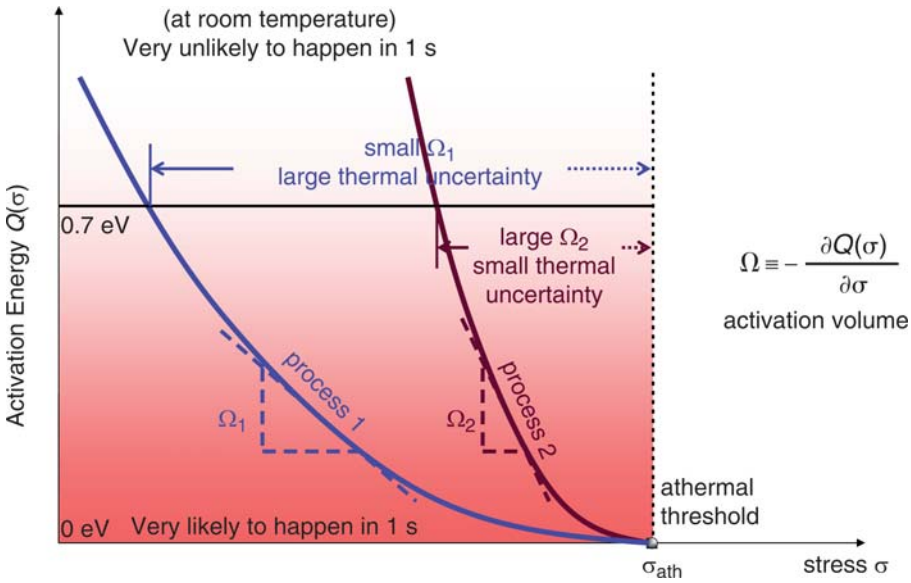

*Figure 2. Illustration of activation energy Q(σ) of stress-driven activated processes. Two processes with the same athermal threshold stress σ_{ath} but different activation volumes are compared. A larger activation volume means the Q curve is steeper, the activated process is more "collective," and correspondingly there is less thermal uncertainty in the stress at which it might happen, indicated by the range of the horizontal arrows. For instance, point defect migration usually has an activation volume $\Omega \sim 0.02\mathbf{b}^3$–$0.1\mathbf{b}^3$, while forest dislocation cutting usually has $\Omega \sim 10^3\mathbf{b}^3$ (where **b** is the Burgers vector). This means the latter is a much more "athermal" process, unlikely to happen unless the applied stress σ is above 90% of σ_{ath} (for a material with 200 MPa yield strength), while diffusion can happen at almost any stress (although usually at somewhat higher temperature). This large separation in the scale of activation volume is the physical basis for the distinction between yield and creep.*

criterion specifies at what load the free energies of the initial and final states are equal. K_{ath} specifies at what load the activation barrier vanishes, so kinetically the crack can move even at 0 K. The difference between K_c and K_{ath} is called the lattice trapping range.[59,60] Similar considerations hold for all defect processes.

The athermal threshold σ_{ath} is usually the first thing one needs to know regarding any homogeneous or heterogeneous nucleation process, because it sets the limit on the $Q(\sigma)$ characteristics. For instance, σ_{ath} is called the Peierls stress when discussing double kink nucleation on a dislocation. With atomistic calculations, the athermal threshold condition can be precisely determined by normal mode analysis,[70,71] where one looks at softening—that is, vanishing of the frequency—of any vibrational mode. This is because as $\sigma \to \sigma_{ath}$, the initial state and the saddle state on the MEP merge, and the stability domain of the free energy basin vanishes along a particular direction, which is that of the MEP. In systems with translational symmetry, such as crystalline bulk,[72] surfaces/interfaces,[6] or wires,[6] one can further exploit this symmetry to simplify the analysis.[73]

A particularly fortunate circumstance arises when the soft normal mode is a long-wavelength phonon or elastic wave,[72] for which a closed-form expression,

$$\Lambda_{\min}(\sigma) = \min_{|\mathbf{k}|=|\mathbf{w}|=1} (C_{ijkl}k_jk_l)w_iw_k + \sigma_{ji}k_jk_l, \quad (2)$$

can be derived for the athermal threshold: $\Lambda_{\min}(\sigma) = 0$, where C_{ijkl} is the elastic constant at finite stress, **k** is the normalized wave vector, and **w** is the polarization displacement of the trial elastic wave. When $\Lambda_{\min}(\sigma) \to 0$ as $\sigma \to \sigma_{ath}$, the system undergoes instability of essentially the same nature as in spinodal decomposition,[74] except the conserved quantity here is not chemical concentration, but the total strain (elastic strain plus inelastic or transformation strain). Instead of two chemical phases precipitating out as in spinodal decomposition, here inelastic strain-carrying defects would "precipitate out," in a matrix of elastic strain-relieved crystal. When $\mathbf{k} \parallel \mathbf{w}$, that is, the soft elastic wave is longitudinal and strain fluctuations are of tensile character, a microcrack is likely to result following the phonon instability. If, on the other hand, $\mathbf{k} \perp \mathbf{w}$, that is, the soft elastic wave is transverse and strain fluctuations are of a shear nature, dislocations are likely to be nucleated. This criterion has been shown in very low-temperature MD simulations of nanoindentation to be capable of predicting the precise location and character (Burgers vector, slip plane) of homogeneously nucleated dislocations.

Not all crystals reach their athermal limit by Equation 2. Indeed, most crystals may not:[70] even simple lattices like fcc Al have non-Brillouin zone center phonons that first become unstable under shear.[75] Nonetheless, Equation 2 may still be a useful indicator because usually an entire phonon branch gets soft together in a particular **k** direction, and whether the zone-center or zone-boundary phonon frequency touches zero first often does not make a big difference in σ_{ath} prediction. Collecting and condensing diffuse elastic strain into localized inelastic strain,[76] causing free energy reduction, is the *raison d'être* of large defects like dislocations and cracks. Since $\Lambda_{\min}(\sigma)$ characterizes the elastic stiffness of the medium, it also comes into play for elastically softened materials at high stress.

Activation Volume and Activation Exponent

With atomistic calculations[54–56,58,60,66] one can compute $Q(\sigma)$. Since one can only perform a finite number of calculations, it is very important to have a physically sound functional form for $Q(\sigma)$ to interpolate the data. The activation volume is defined as $\Omega(\sigma) \equiv -\partial Q/\partial\sigma$. Because σ is a tensor, Ω is generally a strain-like symmetric tensor with unit volume. The intuitive interpretation of Ω is the total volume of coherently activated atoms, multiplied by their transformation strain[77] at the saddle with respect to the initial state. The more "collective" a certain activation event is, the larger its activation volume. Thus, thermally activated forest dislocation cutting tends to have a much larger activation volume ($\sim 10^3\mathbf{b}^3$) than thermally activated vacancy hops ($\sim 0.02\mathbf{b}^3$ to $0.1\mathbf{b}^3$), since the former involves the collective fluctuation of a dislocation segment whose length scale is defined by external pinning points, while vacancy motion is mainly one atom's "decision." The activation volume $\Omega(\sigma)$ is clearly a very important quantity to compute,[37] because once we know the activation energy and activation volume at a reference load $\sigma = \sigma_0$, we can predict the activation energy and therefore nucleation rates at loads not too far from σ_0, by the Taylor expansion

$$Q(\sigma) \approx Q(\sigma_0) - \mathrm{Tr}\Omega(\sigma_0)(\sigma - \sigma_0), \quad (3)$$

where Tr is the trace operation performed on a matrix. In other words, Ω is the tangent slope of the $Q(\sigma)$ curve versus σ if plotted in 1D (see Figure 2) and can be used for extrapolating. Experimentally, $\Omega(\sigma_0)$ is inversely proportional to the strain-rate sensitivity at $\sigma = \sigma_0$ for a macroscopic sample and thus can be measured.[35] It can also be measured for nanoscale systems exhibiting discrete relaxation events by statistical techniques.[23,26]

As a Taylor expansion, Equation 3 cannot cover the whole range of stress accurately. In particular, one cannot perform Taylor expansion at $\sigma = \sigma_{ath}$, because $\Omega(\sigma_{ath}) = 0$. This is to say, $Q(\sigma)$ must approach zero value at $\sigma = \sigma_{ath}$ with zero slope, as illustrated in Figure 2. This is a rigorous result, because by definition, at $\sigma = \sigma_{ath}$ the initial state and the saddle state merge, so the amount of activated transformation must approach zero as $\sigma \to \sigma_{ath}$.

This requirement is satisfied by a "3/2 model," as discussed by Cahn and Nabarro[61] and Cottrell.[62] In this model, $Q(\sigma)$ behaves as $\propto(\sigma_{ath} - \sigma)^{3/2}$ when $\sigma \to \sigma_{ath}$. This would mean the activation volume $\Omega(\sigma)$ drops as $(\sigma_{ath} - \sigma)^{1/2}$ when $\sigma \to \sigma_{ath}$. While the basis for the derivation of the 3/2 model is quite general, it may not be applicable to all thermally activated processes, in particular, those nucleation processes whose rates are limited by bifurcation-type events. To keep the discussion general, we may assume that $Q(\sigma)$ can be fit well by the form $A(\sigma_{ath} - \sigma)^{\alpha}$, where $\alpha > 1$ is called the activation exponent and α,A are independent of stress. The three parameters σ_{ath}, α, and A thus completely specify $Q(\sigma)$ and $\sigma(\sigma)$ functional forms and can be used to interpolate modeling or experimental results.

Cracks and Dislocations

A bond can break in tension or shear.[5] In the latter case, a new bond will usually form immediately afterward between the new atomic neighbors. Cracks and dislocations are bond-breaking "machines" that nature builds to relieve the diffuse elastic strain energy stored elsewhere in the matrix.[76] These machines are very efficient, in the sense that it takes relatively small stress σ_∞ to operate them continuously. This is because cracks and dislocations work like levers, amplifying the far-field stress σ_∞ so that the few bonds inside their cores literally sustain a stress level (in the generalized sense) approaching the ideal strength and, thus, break. It does cost something, sometimes a lot, to build these sophisticated machines. This initial payment is the nucleation barrier.

It is usually beneficial to regard cracks and dislocations as singularities at the continuum scale. But figuring out the absolute formation energy of a true singularity is impossible. With the Peierls model (cohesive zone model for cracks),[51,63] where nonlinear, nonconvex potential energy terms[68] are added, these singularities are regularized, having a finite size,[78] thus allowing activation energy calculations to be carried out.[52] The Peierls framework is absolutely essential for discussing dislocation nucleation: the theory is asymptotically exact in the limit of large dislocation core size. How to calibrate this framework[79] with the activation energy[54–56,58,60,66] and activation volume[37] results from atomistic calculations is a great challenge facing the modeling communities. Some basic issues, such as how to properly define and calculate the absolute dislocation core energy atomistically, have only recently been worked out.[80,81]

Figure 3 shows an example of the calculated MEP of dislocation nucleation from a crack tip in Cu.[54] Sequential atomistic configurations on the activation pathway are rendered in Figure 3a, with the saddle-point configuration shown in Figure 4a. Compared to point-defect processes and most chemical reactions, this reaction is clearly more "collective" (larger activation volume), involving tens to hundreds of "non-trivially activated" atoms with new nearest neighbors at the saddle point, and significant elastic adjustments in the matrix. To make connections with the Peierls framework, we calculate the relative displacement between atoms on two sides of the slip plane. This discrete data set is interpolated to form an inelastic displacement field $\boldsymbol{\delta}(\mathbf{x})$, which is further decomposed into inelastic shear displacement $\boldsymbol{\delta}_{\parallel}(\mathbf{x})$ parallel to the slip plane (Figure 4b), and tensile opening component $\boldsymbol{\delta}_{\perp}(\mathbf{x})$ normal to the slip plane (Figure 4d). The dislocation core is best visualized by looking at $|\nabla\delta_{\parallel}(\mathbf{x})|^2$ (Figure 4c), showing that the core is simply the domain wall between inelastically sheared and unsheared regions. Yet, in the heart of this shear-dominant secondary singularity, there is also a small tensile component. Figure 4d shows that $\boldsymbol{\delta}_{\perp}(\mathbf{x})$ is maximized near where $|\nabla\delta_{\parallel}(\mathbf{x})|^2$ is maximized. Such are the intricacies of shear-tension coupling, with one kind of singularity giving birth to another kind. For example, we know experimentally that when several dislocations are piled up against a hard interface, a microcrack may be nucleated, often leading to fracture initiation. This would then be classified as a shear singularity giving birth to a tensile singularity. This classification scheme can be applied to most heterogeneous defect nucleation processes (Table II).

Defect Processes in Nanoindentation

Despite active investigations of the last decade,[16] our understanding of material response to nanoindentation is still incomplete. While the soft phonon explanation gives a rudimentary account of the high strength observed (Table I), the criterion itself is strictly valid only in the ideal limit of very low temperature and defect-free surface and bulk. Using *in situ* nanoindentation inside a transmission electron microscope, Minor et al. observed dislocation activity prior to major displacement bursts.[15] Yet the maximum shear strength inferred at the major displacement bursts is still very high, ~2 GPa. For comparison, the relaxed $(111)\langle 11\bar{2}\rangle$ ideal shear strength of Al computed from DFT is 2.8 GPa.[5,8] One must note, however, that constraining pressure has a significant effect on the ideal shear strength (Table III). The stress state induced by the indenter is quite complex, since materials are anisotropic and are also softened nonlinearly by the high shear stress (when approaching the top of the Frenkel stress–strain sinusoid) while they are hardened by the confining pressure. Only state-of-the-art stress modeling[82,83] is capable of estimating the true stress distribution under the indenter. In this sense, the popular linear elasticity formula $\tau_{max} \approx 0.31(6PE^{*2}/\pi^3R^2)^{1/3}$ and the view that there exists a constant ideal shear strength value is convenient, but far from quantitatively accurate.[82,83]

The *in situ* observations[15] point to the possibility of heterogeneous dislocation nucleation at the high-stress spot, aided by preexisting defects, which then trigger the major displacement bursts. These preexisting defects must be sessile—for example, Lomer–Cottrell locks, other sessile dislocation structures,[72] vacancies—because if they were mobile they would move away from the high-stress region very quickly, before heterogeneous nucleation can occur. It would still require very high stress, a significant fraction of the ideal strength, to nucleate mobile dislocations out of these originally sessile structures.[72,84] These heterogeneous nucleation processes have already been observed in our MD simulations of nanoindentation, and they are described and discussed in this context.[72,84] They have been shown to be capable of leading to large displacement excursions (Figure 5).[84] The athermal threshold stress to operate these heterogeneous nucleation sources is typically 1/4 to 1/2 of what it would take if it were homogeneous dislocation nucleation. The heterogeneous dislocation nucleation scenario is certainly possible in view of Reference 15 and the intrinsic uncertainties reflected in Table III.

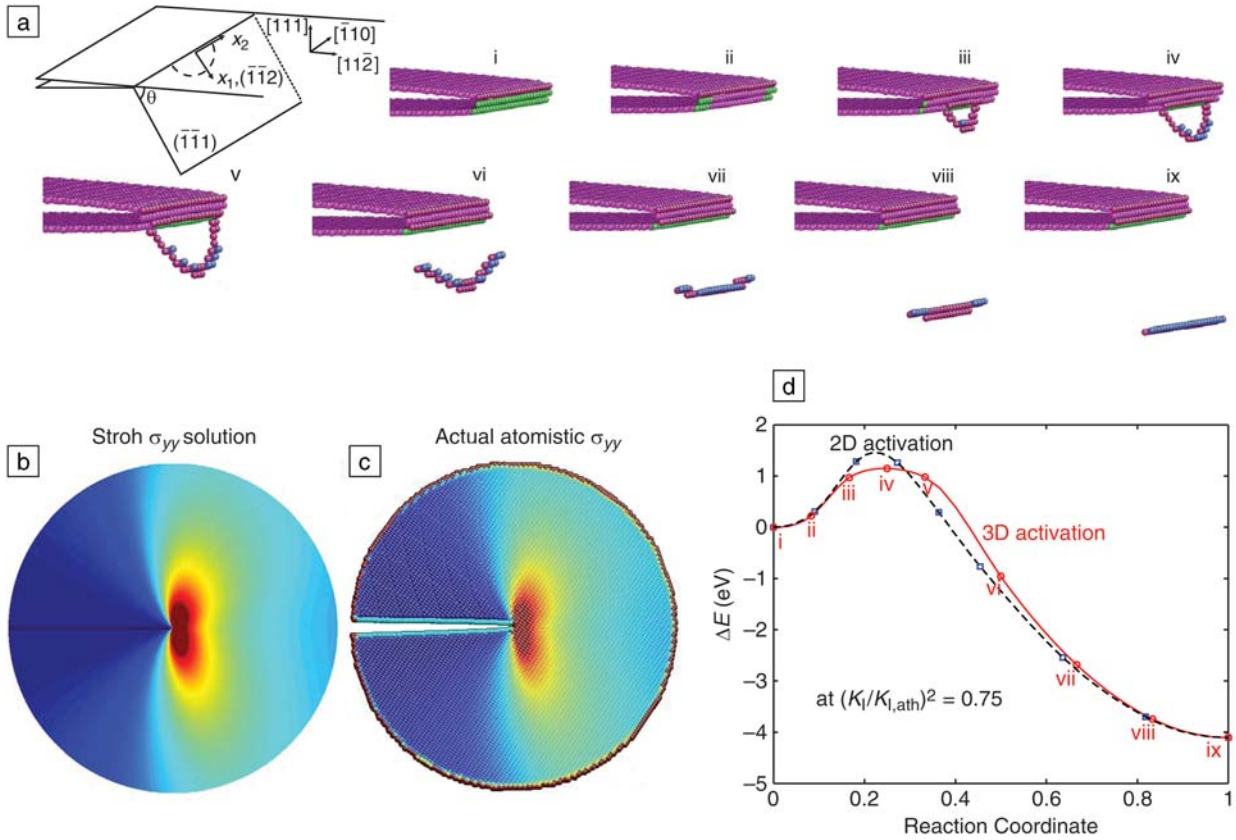

Figure 3. (a) Geometry of a mode I crack containing 24 unit cells (61 Å) in the $\mathbf{x}_2$ *direction and 103,920 Cu atoms in an* R = 80 Å *cylinder.*[54] *Atoms within 5 Å of the cylinder border are fixed according to the anisotropic linear elastic Stroh solution. Schematics (i) to (ix) show the sequential 9 nudged elastic band (NEB) images on the minimum energy path, with (iv) being the saddle point; only atoms whose coordination number differs from 12 are shown. (b) Continuum Stroh solution and (c) the actual atomistic local stress distribution of* σ_{yy} *at* $(K_I/K_{I,ath})^2 = 0.75$. *(d) 3D activation pathway (solid red curve) of partial dislocation emission by bow-out, and its competing 2D pathway, where the dislocation remains straight during emission (dashed curve). Adapted from* Phys. Rev. Lett. ***93*** *025503 (2004).*

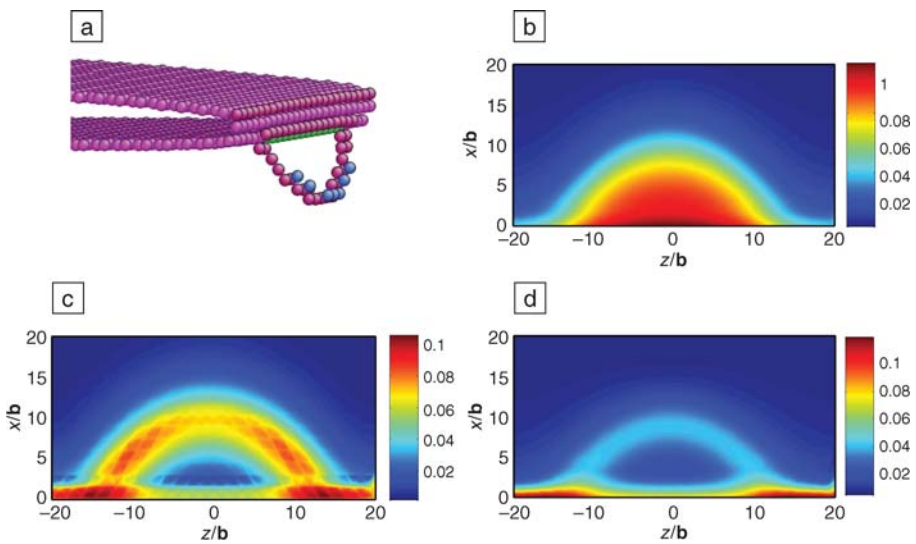

Figure 4. Analysis of the inelastic displacement field $\boldsymbol{\delta}(\mathbf{x})$ *on the inclined slip plane at the saddle point (iv) (see Figure 3a), obtained by spline interpolation of the discrete atomic displacements. (a) Atomic view of the embryonic dislocation loop. (b) Inelastic shear displacement* $\boldsymbol{\delta}_{\parallel}(\mathbf{x})$ *normalized by the partial Burgers vector* $\mathbf{b}_p = a_0[112]/6$, *and (c)* $|\nabla\boldsymbol{\delta}_{\parallel}(\mathbf{x})|^2$. *(d) Tensile opening displacement* $\boldsymbol{\delta}_{\perp}(\mathbf{x})$ *normalized by interplanar spacing* $h_0 = 3^{-1/2}a_0$.

Table II: Examples of Thermally Activated Heterogeneous Defect Nucleation Processes.

Tensile singularity giving birth to tensile singularity:	**Tensile singularity giving birth to shear singularity:**
Cleavage crack kink nucleation[60] Crack deflection at interfaces Mother–daughter crack	Crack-tip dislocation emission[54] Crack-tip twinning[67] Void-induced shear localization
Shear singularity giving birth to tensile singularity:	**Shear singularity giving birth to shear singularity:**
Cracking of brittle inclusions and cracking at grain or phase boundaries in front of dislocation pile-up Crack or void nucleation at dislocation junction or twin intersections	Dislocation cross-slip[55] Dislocation kink nucleation[56,58] Slip transfer reaction (reactants/products include lattice and/or interfacial dislocations) at grain or phase boundaries[37]

Table III: Density Functional Theory Calculated (111)⟨11$\bar{2}$⟩ Ideal Shear Strength of Al and Cu as a Function of Confining Pressure.

	Al (GPa)	Cu (GPa)
P = 0(GPa)	2.84	2.16
P = 10(GPa)	4.49	2.46
P = 20(GPa)	5.90	2.84
σ_{yy} = –10(GPa)	1.78	3.12
σ_{yy} = –20(GPa)	1.41	3.54
σ_{zz} = –3(GPa)	3.64	2.03
$\sigma_{zz} = \sigma_{yy}$ = –10(GPa)	3.98	4.38
$\sigma_{zz} = \sigma_{yy}$ = –20(GPa)	5.26	6.52

P is hydrostatic pressure; σ_{yy} and σ_{zz} is normal stress in the [1$\bar{1}$0] and [111] direction, respectively. Adapted from *Science* **298** (2002) p. 807.

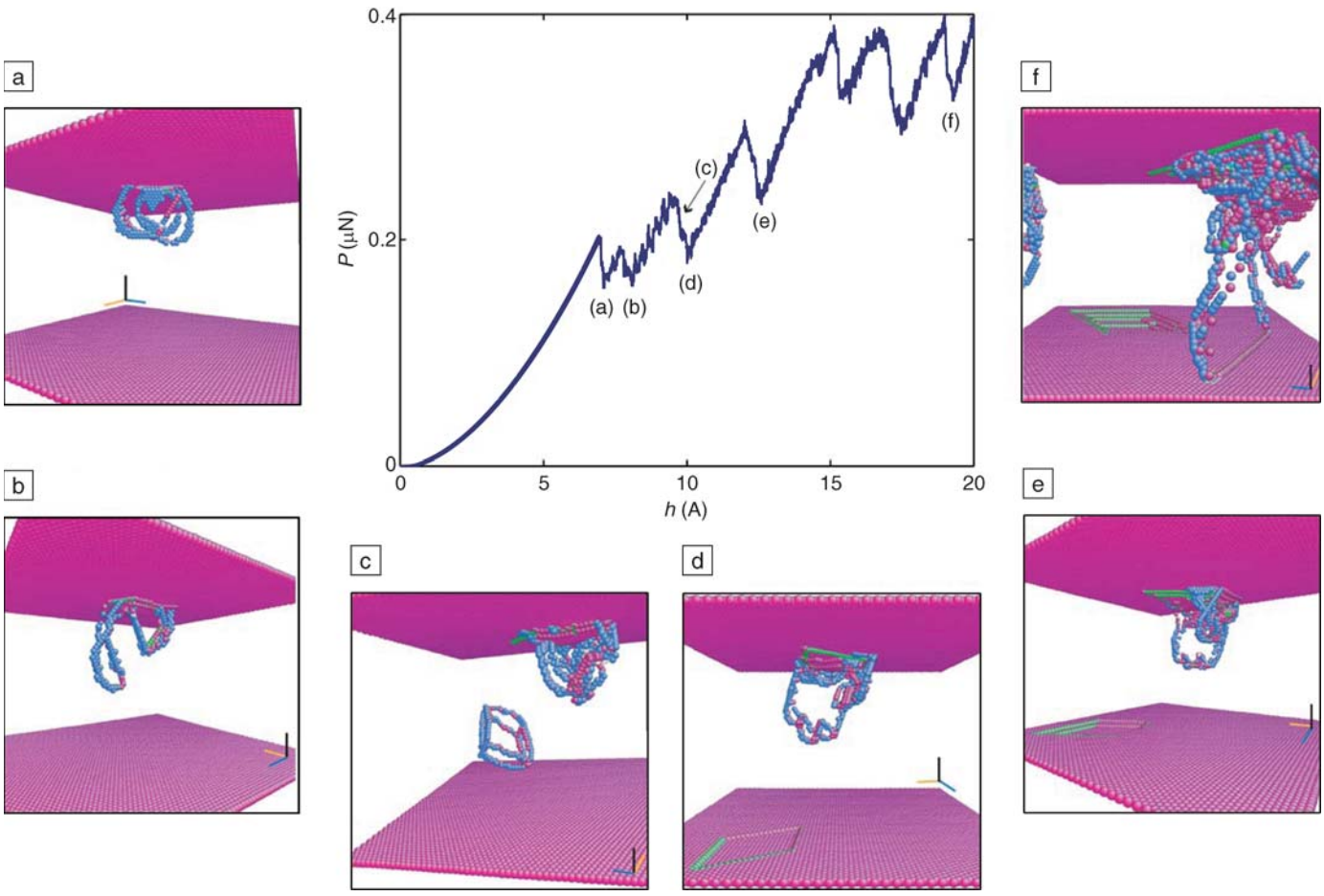

*Figure 5. Molecular dynamics simulation of spherical indentation of {111} single-crystal Al thin film. (a)–(f) Key transitions in the load-displacement (*P–h*) response, with snapshots of atomic activities within the crystal (only atoms whose coordination number differs from 12 are shown). Displacement excursions from (b) to (f) are triggered by heterogeneous nucleation of prismatic dislocation loops and threading dislocations. Adapted from* Phys. Rev. B ***67*** *104105 (2003).*

Using temperature-controlled nanoindentation, Schuh et al.[23] measured the nominal activation volume in Pt to be $\sim 0.5\mathbf{b}^3$. Ngan et al.[26] measured the nominal activation volume in Ni_3Al to be $\sim 1.5\mathbf{b}^3$. These data are intriguing because dislocation processes in conventional size-scale materials tend to have much larger activation volumes. However, our recent atomistic calculations on slip-transfer reactions in nano-twinned Cu[37] and on vacancy-aided and surface-aided dislocation nucleation indicate that Ω can range anywhere from $2\mathbf{b}^3$ to $30\mathbf{b}^3$ at stress σ_0, which gives a very reasonable activation energy $Q(\sigma_0)$ such as 0.5 eV (corresponding to experimental strain rate at room temperature) in these high-strength systems. So dislocation nucleation is not excluded by these experiments. Moreover, there is a distinction between what I call "nominal" and "true" activation volume in nanoindentation. As I mentioned, *linear* elasticity gives a nominal stress prediction σ^* at the high-stress spot, but it may not be what the true stress σ actually is due to nonlinear effects.[82,83] One may, however, define the nominal activation volume to be $\Omega^* \equiv -Q/\sigma^*$, a quantity that is convenient in handling experimental data. Clearly, the nominal activation volume Ω^* is related to the true activation volume by

$$\Omega^* \equiv -Q/\sigma^* = (-Q/\sigma)(\sigma/\sigma^*) = \Omega(\sigma/\sigma^*). \quad (4)$$

One may show, however, that the σ/σ^* factor can be significantly smaller than 1, due to nonlinear elasticity effects. This is in addition to the fact that the true activation volume Ω itself is tending to zero as $\sigma \to \sigma_{ath}$. Combining these factors, calculations show that even homogeneous dislocation nucleation can be surprisingly "thermal," with non-negligible statistical uncertainties in the critical indenter load at nucleation. Whether homogeneous nucleation occurs or not in practice, however, is a function of the perfectness of the sampled volume and loading history.[17,85]

Summary

Defect nucleation is an important problem, linking mechanics, physics, and chemistry at the nanoscale. I have outlined a few threads of this vast subject. Many aspects, such as diffusive (instead of displacive) nucleation processes, are unfortunately not included. Atomistic and first-principles modeling methods are uniquely suited for studying nucleation mechanisms. The present experimental focus on high-strength systems and phenomena provides ample opportunities for such investigation.

Acknowledgments

I would like to thank my mentor Sidney Yip and colleague Ting Zhu, with whom I did much of the work presented here. Special thanks to Ali Argon and Subra Suresh, as well as collaborators Krystyn Van Vliet, Shigenobu Ogata, Yunzhi Wang, Mike Mills, Peter Anderson, Hamish Fraser, Peter Gumbsch, Sergey Dmitriev, and Wei Cai. Support by NSF NANOMESO, ONR D3D, AFOSR MEANS2, DoE, and Honda is gratefully acknowledged.

References

1. M.D. Uchic, D.M. Dimiduk, J.N. Florando, and W.D. Nix, *Science* **305** (2004) p. 986.
2. J.R. Greer, W.C. Oliver, and W.D. Nix, *Acta Mater.* **53** (2005) p. 1821.
3. C.A. Volkert and E.T. Lilleodden, *Philos. Mag.* **86** (2006) p. 5567.
4. W. Kohn, A.D. Becke, and R.G. Parr, *J. Phys. Chem.* **100** (1996) p. 12974.
5. S. Ogata, J. Li, N. Hirosaki, Y. Shibutani, and S. Yip, *Phys. Rev. B* **70** 104104 (2004).
6. S.V. Dmitriev, T. Kitamura, J. Li, Y. Umeno, K. Yashiro, and N. Yoshikawa, *Acta Mater.* **53** (2005) p. 1215.
7. Y. Umeno, A. Kushima, T. Kitamura, P. Gumbsch, and J. Li, *Phys. Rev. B* **72** 165431 (2005).
8. S. Ogata, J. Li, and S. Yip, *Science* **298** (2002) p. 807.
9. M.S. Daw and M.I. Baskes, *Phys. Rev. B* **29** (1984) p. 6443.
10. J. Biener, A.M. Hodge, A.V. Hamza, L.M. Hsiung, and J.H. Satcher, *J. Appl. Phys.* **97** 024301 (2005).
11. C.A. Volkert, E.T. Lilleodden, D. Kramer, and J. Weissmuller, *Appl. Phys. Lett.* **89** 061920 (2006).
12. W.C. Oliver and G.M. Pharr, *J. Mater. Res.* **7** (1992) p. 1564.
13. W.W. Gerberich, S.K. Venkataraman, H. Huang, S.E. Harvey, and D.L. Kohlstedt, *Acta Metall. Mater.* **43** (1995) p. 1569.
14. A. Gouldstone, H.J. Koh, K.Y. Zeng, A.E. Giannakopoulos, and S. Suresh, *Acta Mater.* **48** (2000) p. 2277.
15. A.M. Minor, S.A.S. Asif, Z. Shan, E.A. Stach, E. Cyrankowski, T.J. Wyrobek, and O.L. Warren, *Nature Mater.* **5** (2006) p. 697.
16. A. Gouldstone, N. Chollacoop, M. Dao, J. Li, A.M. Minor, and Y.-L. Shen, *Acta Mater.* (2007) overview no. 142.
17. A. Gouldstone, K.J. Van Vliet, and S. Suresh, *Nature* **411** (2001) p. 656.
18. W.W. Gerberich, J.C. Nelson, E.T. Lilleodden, P. Anderson, and J.T. Wyrobek, *Acta Mater.* **44** (1996) p. 3585.
19. H.D. Espinosa, S. Berbenni, M. Panico, and K.W. Schwarz, *Proc. Natl. Acad. Sci. USA* **102** (2005) p. 16933.
20. J.R. Greer and W.D. Nix, *Phys. Rev. B* **73** 245410 (2006).
21. K. Sieradzki, A. Rinaldi, C. Friesen, and P. Peraltai, *Acta Mater.* **54** (2006) p. 4533.
22. S. Yip, *Nature* **391** (1998) p. 532.
23. J.K. Mason, A.C. Lund, and C.A. Schuh, *Phys. Rev. B* **73** 054102 (2006).
24. J. Pokluda, M. Cerny, P. Sandera, and M. Sob, *J. Comput. Aided Mater. Des.* **11** (2004) p. 1.
25. W. Wang and K. Lu, *Philos. Mag.* **86** (2006) p. 5309.
26. P.C. Wo, L. Zuo, and A.H.W. Ngan, *J. Mater. Res.* **20** (2005) p. 489.
27. A. Asenjo, M. Jaafar, E. Carrasco, and J.M. Rojo, *Phys. Rev. B* **73** 075431 (2006).
28. D. Lorenz, A. Zeckzer, U. Hilpert, P. Grau, H. Johansen, and H.S. Leipner, *Phys. Rev. B* **67** 172101 (2003).
29. C.A. Schuh and A.C. Lund, *J. Mater. Res.* **19** (2004) p. 2152.
30. H. Bei, E.P. George, J.L. Hay, and G.M. Pharr, *Phys. Rev. Lett.* **95** 045501 (2005).
31. H.S. Leipner, D. Lorenz, A. Zeckzer, and P. Grau, *Phys. Status Solidi A* **183** (2001) p. R4.
32. U.F. Kocks, A.S. Argon, and M.F. Ashby, *Prog. Mater. Sci.* **19** (1975) p. 1.
33. K.S. Kumar, H. Van Swygenhoven, and S. Suresh, *Acta Mater.* **51** (2003) p. 5743.
34. L. Lu, Y.F. Shen, X.H. Chen, L.H. Qian, and K. Lu, *Science* **304** (2004) p. 422.
35. L. Lu, R. Schwaiger, Z.W. Shan, M. Dao, K. Lu, and S. Suresh, *Acta Mater.* **53** (2005) p. 2169.
36. R.J. Asaro and S. Suresh, *Acta Mater.* **53** (2005) p. 3369.
37. T. Zhu, J. Li, A. Samanta, H.G. Kim, and S. Suresh, *Proc. Natl. Acad. Sci. USA* (2007) in press, www.pnas.org/cgi/doi/10.1073/pnas.0611097104.
38. Y.M. Wang, M.W. Chen, F.H. Zhou, and E. Ma, *Nature* **419** (2002) p. 912.
39. Y.T.T. Zhu and X.Z. Liao, *Nature Mater.* **3** (2004) p. 351.
40. W.L. Johnson, *MRS Bulletin* **24** (10) (1999) p. 42.
41. F. Shimizu, S. Ogata, and J. Li, *Acta Mater.* **54** (2006) p. 4293.
42. E.M. Bringa, A. Caro, Y.M. Wang, M. Victoria, J.M. McNaney, B.A. Remington, R.F. Smith, B.R. Torralva, and H. Van Swygenhoven, *Science* **309** (2005) p. 1838.
43. S.E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z. Ma, B. McIntyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, and Y. El-Mansy, *IEEE Electron Dev. Lett.* **25** (2004) p. 191.
44. P.R. Chidambaram, C. Bowen, S. Chakravarthi, C. Machala, and R. Wise, *IEEE Trans. Electron Dev.* **53** (2006) p. 944.
45. Z. Zhang, J. Yoon, and Z.G. Suo, *Appl. Phys. Lett.* **89** 261912 (2006).
46. T. Dumitrica, M. Hua, and B.I. Yakobson, *Proc. Natl. Acad. Sci. USA* **103** (2006) p. 6105.
47. S.L. Zhang, S.L. Mielke, R. Khare, D. Troya, R.S. Ruoff, G.C. Schatz, and T. Belytschko, *Phys. Rev. B* **71** 115403 (2005).
48. J.Y. Huang, S. Chen, Z.F. Ren, Z.Q. Wang, D.Z. Wang, M. Vaziri, Z. Suo, G. Chen, and M.S. Dresselhaus, *Phys. Rev. Lett.* **97** 075501 (2006).
49. H. Mori, S. Ogata, J. Li, S. Akita, and Y. Nakayama, *Phys. Rev. B* **74** 165418 (2006).
50. J.R. Rice and R. Thomson, *Philos. Mag.* **29** (1974) p. 73.
51. J.R. Rice and G.E. Beltz, *J. Mech. Phys. Solids* **42** (1994) p. 333.
52. G. Xu, A.S. Argon, and M. Oritz, *Philos. Mag. A* **75** (1997) p. 341.
53. A.S. Argon, *J. Eng. Mater. Technol.-Trans. ASME* **123** (2001) p. 1.
54. T. Zhu, J. Li, and S. Yip, *Phys. Rev. Lett.* **93** 025503 (2004).

55. T. Vegge, T. Rasmussen, T. Leffers, O.B. Pedersen, and K.W. Jacobsen, *Phys. Rev. Lett.* **85** (2000) p. 3866.
56. V.V. Bulatov, S. Yip, and A.S. Argon, *Philos. Mag. A* **72** (1995) p. 453.
57. W. Cai, V.V. Bulatov, J.F. Justo, A.S. Argon, and S. Yip, *Phys. Rev. Lett.* **84** (2000) p. 3346.
58. M. Wen and A.H.W. Ngan, *Acta Mater.* **48** (2000) p. 4255.
59. B.R. Lawn, D.H. Roach, and R.M. Thomson, *J. Mater. Sci.* **22** (1987) p. 4036.
60. T. Zhu, J. Li, and S. Yip, *Phys. Rev. Lett.* **93** 205504 (2004).
61. J.W. Cahn and F.R.N. Nabarro, *Philos. Mag. A* **81** (2001) p. 1409.
62. A.H. Cottrell, *Philos. Mag. Lett.* **82** (2002) p. 65.
63. J.R. Rice, *J. Mech. Phys. Solids* **40** (1992) p. 239.
64. J. Li, A.H.W. Ngan, and P. Gumbsch, *Acta Mater.* **51** (2003) p. 5711.
65. G. Henkelman and H. Jonsson, *J. Chem. Phys.* **113** (2000) p. 9978.
66. T. Zhu, J. Li, X. Lin, and S. Yip, *J. Mech. Phys. Solids* **53** (2005) p. 1597.
67. E.B. Tadmor and S. Hai, *J. Mech. Phys. Solids* **51** (2003) p. 765.
68. V. Vitek, *Scripta Metall.* **4** (1970) p. 725.
69. A. van de Walle, M. Asta, and G. Ceder, *Calphad* **26** (2002) p. 539.
70. J. Li and S. Yip, *CMES-Comp. Model. Eng. Sci.* **3** (2002) p. 219.
71. T. Kitamura, Y. Umeno, and R. Fushino, *Mater. Sci. Eng. A* **379** (2004) p. 229.
72. J. Li, K.J. Van Vliet, T. Zhu, S. Yip, and S. Suresh, *Nature* **418** (2002) p. 307.
73. N. Binggeli, N.R. Keskar, and J.R. Chelikowsky, *Phys. Rev. B* **49** (1994) p. 3075.
74. J.W. Cahn, *Acta Metall.* **9** (1961) p. 795.
75. D.M. Clatterbuck, C.R. Krenn, M.L. Cohen, and J.W. Morris, *Phys. Rev. Lett.* **91** 135501 (2003).
76. S. Ogata, J. Li, and S. Yip, *Phys. Rev. B* **71** 224102 (2005).
77. A.G. Khachaturyan, *Theory of Structural Transformation in Solids* (Wiley, New York, 1983).
78. R. Peierls, *Proc. Phys. Soc. London* **52** (1940) p. 34.
79. V.V. Bulatov and E. Kaxiras, *Phys. Rev. Lett.* **78** (1997) p. 4221.
80. W. Cai, V.V. Bulatov, J.P. Chang, J. Li, and S. Yip, *Phys. Rev. Lett.* **86** (2001) p. 5727.
81. J. Li, C.Z. Wang, J.P. Chang, W. Cai, V.V. Bulatov, K.M. Ho, and S. Yip, *Phys. Rev. B* **70** 104113 (2004).
82. T. Zhu, J. Li, K.J. Van Vliet, S. Ogata, S. Yip, and S. Suresh, *J. Mech. Phys. Solids* **52** (2004) p. 691.
83. R.L. Hayes, M. Fago, M. Ortiz, and E.A. Carter, *Multiscale Model. Simul.* **4** (2005) p. 359.
84. K.J. Van Vliet, J. Li, T. Zhu, S. Yip, and S. Suresh, *Phys. Rev. B* **67** 104105 (2003).
85. P. Schall, I. Cohen, D.A. Weitz, and F. Spaepen, *Nature* **440** (2006) p. 319.

Ju Li *is an assistant professor in the Department of Materials Science and Engineering at the Ohio State University. He received his bachelor's degree in physics from the University of Science and Technology of China in 1994, and his PhD degree in nuclear engineering from the Massachusetts Institute of Technology in 2000. He was a postdoctoral associate and a research scientist at MIT before joining the faculty at Ohio State in the fall of 2002.*

Li's research focus is the development of robust analytical and computational approaches to the modeling of the structural and functional properties of materials, including Ni- and Ti-based superalloys, fuel cell catalysts, ultrahigh-temperature ceramics, hydrogen storage materials, electroactive polymers, and metallic glasses. In these endeavors, Li is leading the rapidly growing multiscale modeling effort in bridging continuum, atomistic, and electronic scales at the interface of nanoscience and mechanics.

He has more than 60 peer-reviewed publications and is the author of the molecular visualization software AtomEye. Li won an MRS Graduate Student Silver Award in 1999 and was the recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE) 2005.

Li can be reached by e-mail at li.562@osu.edu.